J2
409/2:3
c.2
NORTH CAROLINA STATE LIBRARY
RALEIGH
N. C.
Doc.
JUN 15 1978

PHSB STUDIES

A Special Report Series by the N.C. Department of Human Resources, Division of Health Services, Public Health Statistics Branch, P.O. Box 2091, Raleigh, N.C.

No. 3 April 1977

ASSOCIATIONS BETWEEN MORTALITY AND VARIOUS SOCIAL, ECONOMIC AND ENVIRONMENTAL FACTORS IN NORTH CAROLINA

(An Indirect Study)

For nearly two decades, the question of a relationship between health and the mineral content of water has been cause for epidemiologic studies both in the United States and abroad. Although these studies have failed to produce conclusive evidence as to the identity of any underlying "water factor," there is a general impression that one or more such entities exist and that cardiovascular disease is a particularly apt effect. To date, the hypothesis most often formulated has been that of a negative association between cardiovascular disease and water hardness. (1)

In an attempt to elucidate the water story in North Carolina, Brittain (2) has studied water hardness while Voors (3) has assessed calcium and potassium in drinking water as possible determinants of heart disease mortality. Findings for hardness were not in the expected direction, but Voors found—for inner plain and eastern Piedmont counties—that water calcium correlated negatively and water potassium positively with "sudden" atherosclerotic heart disease mortality among white males.

There is also mounting evidence that higher elevation favors improved heart function (4,5). With its physical gradients ranging in elevation from sea level to peaks of more than 6,000 feet, North Carolina seems most suitable to investigations of the altitude question.

The present study is a further attempt to elucidate North Carolina's water story while examining elevation and other possible correlates of cause-specific mortality. For eighty-one of the State's 100 counties, multiple linear regression is used to determine what factors, including 15 water constituents and 17 other environmental and socioeconomic indicators, might be affecting selected age-race-sex-adjusted death rates. In these analyses, occupational distribution is used to control for occupation-related factors not otherwise accounted for in regression.

METHODS AND MATERIALS

Mortality Data

Rates used in these analyses are the 1973-75 average annual death rates, adjusted for age, race and sex (6) and specific for the four disease entities described on the next page. Other causes of death were not studied due to the fact that death rates are presently computed only for underlying causes of death, and for many diseases, the "incidence at death" is known to be much higher. For example, hypertension, arteriosclerosis and diabetes are considered "associated" conditions far more often than they are considered an underlying cause. Since our primary interest is in the determinants of disease rather than the determinants of underlying causes of death per se, analysis based solely on underlying frequencies would not be appropriate. We also did not study causes for which 3-year rates appear to fluctuate randomly over time.

The following underlying causes were deemed amenable to study on the basis that (i) the underlying frequencies are thought to closely represent the incidence at death and (ii) correlations between the 1968-70 and 1973-75 county death rates (100 counties) are statistically significant suggesting the existence of some local condition that may invite intervention.

Cause	ICDA Codes (7)
Acute Myocardial Infarction	410
Colon and Rectum Cancer	153,154
Trachea, Bronchus and Lung Cancer	162
Prostatic Cancer	185

Water Supply Data

The Laboratory Section of the Division of Health Services analyzed 1,860 water supply samples during 1973. Of these, 89% were from supplies that each served ten or more residential connections and together served an estimated 57 percent of the State's 1973 population (8). The remaining samples were from industrial, institutional, recreational and private supplies.

Eighty-seven percent of the samples were ground water. However, surface water is used by many of the larger supplies and accounted for more than three-quarters of the public water supply population.

Approximately 70 percent of the samples were raw as opposed to treated water. Sources of the samples were: well tap (42%), tap in the distribution system (34%) and house tap (24%). Sources in the Division's Water Supply Branch indicate that type (treated or raw) and tap location are such that the samples obtained closely represent the drinking water of the population served.

Among the 1,653 supplies serving residential connections, most water constituent values were within established limits (9). Statewide, the supplies were characterized by a preponderance of soft water (about 90 percent of the population served), a high percentage of turbid water (about 40 percent of the population served), and a high percentage of low-fluoride water (about 40 percent of the population served). Iron and manganese levels were excessive in water served to approximately 5 and 8 percent respectively of the population. Much lower percentages of the population used water containing excessive amounts of other minerals.

Calculation of County Water Constituent Values

The algorithm used to compute a county's average water constituent values is similar to that employed by Brittain (2) and involves two components: (i) a component attributable to the population served by reporting public supplies and (ii) an estimate of the component attributable to the population served by unreported and private supplies. The county's final average constituent value is the population-weighted average of these two components.

Formulas used in the calculations are available from the Public Health Statistics Branch. Briefly, the first component is the population-weighted average value for those county water supplies serving residential connections while the second component is the unweighted average value for all ground water supplies. The latter component assumes that the unreported and private sectors of a county use ground more than surface water and that all ground water samples are equally likely to be representative. Finally, the two components are population-weighted and combined.

As previously stated, this study includes 81 of the State's 100 counties. Counties eliminated from analysis were those whose reported ground water supplies were considered too few in number to provide an acceptable estimate of the second component required in the preceding algorithm. A county was eliminated if

Percentage Population Served by Reporting Supplies Was		Number of Reported Ground Water Supplies Was Fewer Than
Under 20	AND	5
20-39		4
40-59		3
60-79		2

While these criteria are much less restrictive than desired, they were imposed for the sake of retaining a representative geo-distribution of counties. Even so, seven coastal counties and 5 westernmost counties were among the 19 counties eliminated.

In addition to the exclusion of counties, we also excluded values for cadmium, arsenic, chromium^{+6} and lead from analysis since these were associated with extremely low variability and hence did not allow for contrasting the effects of diverse values. In this regard, values for hardness were also less diverse than desired for a study of this type.

The fifteen water variables exposed to analysis are: ph, turbidity, alkalinity, acidity, hardness, calcium, magnesium, iron, manganese, sodium, chloride, potassium, fluoride, copper and zinc.

Occupation Data

The percentages of the labor force employed in 33 occupation categories at the time of the 1970 Census are used to control for various occupation-related factors not otherwise accounted for in regression.

Other Socio-economic and Environmental Data

Derived from various sources, the 17 variables are as follows:

- *Per capita income: 1970 Census*
- *Median years of school for persons over age 25: 1970 Census*
- *Dependency load index (ratio of under 18 and 65+ population to population 18-64 years): 1970 Census*
- *Marital unrest index (ratio of divorced and separated persons to married persons): 1970 Census*
- *Population density (population per 1,000 land acres): 1973*
- *Population per primary care physician: 1973*
- *Elevation of county seat (feet)*
- *Precipitation (inches): 30-year average, 1931-1960*
- *Mean January temperature: 30-year average, 1931-1960*
- *Mean July temperature: 30-year average, 1931-1960*

- *Cotton Acreage (harvested acres per 1,000 land acres): 1974*[@]
- *Tobacco Acreage (harvested acres per 1,000 land acres): 1974*[@]
- *Air particulates (tons per 1,000 land acres): 1972*
- *Air sulfur dioxide (tons per 1,000 land acres): 1972*
- *Air carbon monoxide (tons per 1,000 land acres): 1972*
- *Air hydro-carbon (tons per 1,000 land acres): 1972*
- *Air nitrogen dioxide (tons per 1,000 land acres): 1972*

Regression Analysis

In order to assess the independent effects of the various socio-economic and environmental factors (the predictor variables) on each of the death rates (the outcome variables), we have employed stepwise multiple regression (10,11). In so doing, we were forced in some instances to allow a single predictor variable to serve as proxy for a set of predictor variables. For example, we could not use both hardness and calcium in a single regression since they are almost perfectly correlated; we arbitrarily chose hardness as the proxy variable.

Table 1 shows pairs of intercorrelated variables. In each case, Variable X served as the proxy variable in regression, and we presume that results for that variable imply results for its correlates, either directly or inversely depending on the direction of Table 1 correlations. Although this presumption may not be sound in all cases, the alternative (i.e., to test all possible models) was considered economically unfeasible.

In addition to the regression model reported here, we did test acute myocardial infarction against a number of alternate models. For instance, we examined a model that contained no occupation variables and several models based on different sets of occupational classifications including a reduced set produced statistically. It is noteworthy that the model with no occupational variables explained about 27 percent of the variation in the death rate whereas the model with the optimal selection of variables explained around 74 percent. Our conclusion was that the detailed occupation variables presently used are highly important and should not be excluded.

TABLE 1

Pairs of Intercorrelated Variables*

81 North Carolina Counties

Variable X	Variable Y	r_{xy}
Population Density	Air Hydro-carbon	.83
Population Density	Air Carbon Monoxide	.93
Air Particulates	Air Sulfur Dioxide	.91
Air Particulates	Air Nitrogen Dioxide	.84
Elevation	Mean January Temperature	-.84
Elevation	Mean July Temperature	-.92
Water Sodium	Water Alkalinity	.90
Water Sodium	Water Chloride	.79
Water Sodium	Water Potassium	.84
Water Hardness	Water Calcium	.98
Water Hardness	Water Magnesium	.79

*Absolute value of Pearson's product-moment correlation coefficient (r) is approximately .80.

[@] Included as an indicator of pesticide usage.

RESULTS

Among the death rates examined, all were well explained by regression. The coefficients of determination (R^2) ranged between .74 and .83 meaning that regression explained an estimated 74 to 83 percent of the variation in the respective death rates. Each of these estimates was statistically significant.

As indicated in the preceding section, occupational distribution was found to be highly explanatory for acute myocardial infarction. The occupation variables also made significant contributions to the explanation of lung and prostatic cancer but were much less explanatory for colon-rectum cancer. Due to the large number of variables as well as difficulty of interpretation, specific results for the occupation variables are not given here.

For other variables used in these analyses, Table 2 summarizes the statistically significant results. Here, parentheses ( ) are used to identify a set of highly correlated predictor variables; the symbol † identifies the proxy variable used in regression. In synthesizing these results, one should keep in mind that results for a proxy variable are probably applicable to that variable's correlates, either directly or inversely, depending upon the direction (sign) of the correlation between the proxy variable and the correlate (see Table 1).

It is also important to keep in mind that these results are subject to the usual limitations of regression analysis; for example, results are biased to the extent that potentially causative variables (such as nutritional factors and alcohol and tobacco consumption) could not be included and have not been accounted for.

Finally, one should keep in mind that these results, while taking a large number of factors into account, in no way imply cause-and-effect relationships. They merely indicate the extent to which a death rate is associated with a certain predictor variable, given that we have simultaneously accounted for associations between the death rate and other predictor variables being considered.

TABLE 2

Statistically Significant Results for Non-occupation Variables
Regressed on Age-race-sex-adjusted Death Rates

81 North Carolina Counties

	Statistically Significant Variables		
Cause of Death	Name	Beta Weight*	Level of Significance**
Acute Myocardial Infarction	(Elevation†, January Temperature, July Temperature)	-1.054	.005
	Water Zinc	.656	.01
	Population per Physician	-.585	.05
Lung Cancer	Per Capita Income	-2.052	.005
	Cotton Acreage	-.587	.01
	Population per Physician	-.485	.05
	Water Iron	-.350	.05
Colon-rectum Cancer	Median Education	.860	.025
	Water Manganese	.509	.005
Prostatic Cancer	(Elevation†, January Temperature, July Temperature)	-.740	.025
	Precipitation	-.553	.01
	(Water Sodium†, Alkalinity, Chloride, Potassium)	-.530	.025
	(Air Particulates†, Sulfur Dioxide, Nitrogen Dioxide)	-.364	.025

*Absolute value may be used to compare strength of 2 or more variables within a single regression. Sign shows direction of association with the death rate.

**The lower the number, the greater confidence one may have in the precision of the beta weight.

†A proxy variable, i.e., results for this variable imply results for its correlates (see explanation, p. 4).

STATE LIBRARY OF NORTH CAROLINA
3 3091 00739 0297

Given the preceding considerations, highlights of this study and some related commentary follow:

- *Considered against other variables used in this study, water hardness was not predictive of death from acute myocardial infarction. However, as indicated earlier, values for hardness may not be sufficiently diverse to adequately assess the hardness factor.*
- *Although water zinc was found to be a positive correlate of myocardial infarction, it is noted that levels of zinc in North Carolina appear well within established limits. Still, perhaps some adverse effect of zinc upon the cardiovascular system should not be ruled out; and alternatively, one might consider the prospect of "competing risks" where zinc may be protective against some other disease entities thereby increasing the risk of death from myocardial infarction.*
- *Irrespective of model specification, inaccessibility of primary care physicians (as measured by the population per physician ratio) was shown to be negatively associated with death from myocardial infarction. While cause-and-effect presumably is not involved in this relationship, it seems clear that accessibility of physicians is not a major deterrent to death from this cause.*
- *Elevation was by far the strongest predictor of acute myocardial infarction. A negative correlate, this predictor implies that temperature, especially summer temperature, is a positive correlate.*
- *Accessibility of primary care physicians also would not appear a deterrent to lung cancer mortality. At the same time, per capita income, cotton acreage and water iron were negative correlates, per capita income being by far the strongest predictor.*
- *Median education and water manganese were positively associated with death from colon-rectum cancer. Contrary to the case for zinc, high manganese levels are found in North Carolina.*
- *Prostatic cancer mortality was found to be negatively associated with elevation, precipitation, air particulates and its correlates, and water sodium and its correlates. As was true for acute myocardial infarction, temperature may be considered a positive correlate.*

Although it is entirely possible that any or all of the above associations are spurious, at least one finding is not without support; that is, the finding of a negative association between elevation and acute myocardial infarction. Not only is such an association superficially evident in North Carolina, but various studies recently reviewed by Mortimer (4) largely support the hypothesis of a protective effect of altitude. These studies have involved indirect as well as physiologic evidence, the latter including rat experiments in which survival of induced myocardial infarction was significantly greater in the presence of high altitude acclimatization (5).

At the same time, the association presently observed may well reflect climatic stress as well as elements of life style not accounted for by occupation, income, education, etc. In this regard, diet is being increasingly investigated for its role in disease, and data from the North Carolina Nutrition Survey indicate that in 1970 dietary practices of highland North Carolinians were better than those observed in the east (12,13). Hence, elevation per se may not be the protective factor in this state.

With this in mind, a subsequent report examines associations between mortality and nutrition factors in the state's six health service areas since, unfortunately, dietary data are not available for counties.

A final word is that this study was undertaken for two reasons, first, to contribute whatever possible to the identification of factors that are or are not associated with four causes of death, and secondly, to develop methodology to apply to these and other causes as soon as data for multiple conditions at death become available. Thus, comments concerning methodology, data input or any other aspect of this study are invited.

REFERENCES

(1) Neri, L.C., Hewitt, D. and Schreiber, G.B. "Reviews and Commentary: Can Epidemiology Elucidate the Water Story?" American Journal of Epidemiology, Vol. 99, No. 2, pp. 75-86, February 1974.

(2) Brittain, R. "An Ecological Study of the Relationship between Water Hardness and Coronary Heart Disease Mortality in the State of North Carolina." Master's Thesis, Department of Environmental Science and Engineering, School of Public Health, University of North Carolina, Chapel Hill, 1968.

(3) Voors, A.W. "Atherosclerotic Heart Disease and Drinking Water Trace Metals in North Carolina." Trace Substance in Environmental Health V, pp. 523-534, 1972.

(4) Mortimer, E.A., Jr., Monson, R.R., and MacMahon, B. "Reduction in Mortality from Coronary Heart Disease in Men Residing at High Altitude." The New England Journal of Medicine, Vol. 296, No. 11, pp. 581-585, March 17, 1977.

(5) Meerson, F.Z., Gomzakov, O.A., and Shimkovick, M.V. "Adaptation to High Altitude Hypoxia as a Factor Preventing Development of Myocardial Ischemic Necrosis." The American Journal of Cardiology, Vol. 31, pp. 30-34, January 1973.

(6) Department of Human Resources, Division of Health Services, Public Health Statistics Branch. Leading Causes of Mortality, North Carolina Vital Statistics, 1973-75, Vol. 2. Raleigh, July 1976.

(7) U.S. Department of Health, Education, and Welfare. Eighth Revision International Classification of Diseases, Adapted for Use in the United States. Public Health Service Publication Number 1693, U.S. Government Printing Office, Washington, D.C., December 1968.

(8) North Carolina Department of Human Resources, Division of Health Services, Public Health Statistics Branch. North Carolina Vital Statistics 1973, Vol. 1. Raleigh, April 1973.

(9) North Carolina State Board of Health. Chemical Analysis of Water, Form ES-20 5-70. Laboratory Division, Raleigh.

(10) Neter, John and William Wasserman. Applied Linear Statistical Models. Richard D. Irwin, Inc. Homewood, Illinois, pp. 382-385, 1974.

(11) Dixon, W.J., ed. Biomedical Computer Programs, BMDP2R. University of California Press, Berkeley and Los Angeles.

(12) North Carolina Department of Human Resources, Division of Health Services, Administrative Services Section, Public Health Statistics Branch. "Preliminary Regression Analysis of the North Carolina Nutrition Survey Data" by Paul W. C. Johnson. Statistical Report Series (SRS-2210-07-30-73). Raleigh, July 30, 1973.

(13) North Carolina State Board of Health. North Carolina Nutrition Survey, Part One. Raleigh, July 19, 1971.

We are indebted to Dr. Roger Grimson, statistical consultant from the Department of Biostatistics, School of Public Health, University of North Carolina, for assisting in this study.

Public Health Statistics Branch
Division of Health Services
Department of Human Resources
P. O. Box 2091
Raleigh, North Carolina 27602